

DRAFT MEETING RECORD  
Video Compression Measurements Subcommittee G-2.1.6  
Audio Video Techniques Committee G-2.1

Broadcast Technology Society  
Institute of Electrical and Electronics Engineers

Eighth Meeting

NTIA / ITS  
U.S. Department of Commerce  
325 Broadway  
Boulder, CO 80303

March 16, 1998

Item 1 - Welcome and Introduction by Interim Chairman, of IEEE G-2.1.6.

Alan Godber called the meeting to order at 1:09 PM.

Item 2 – Approval of Draft Agenda.

Because some attendees had not reviewed the minutes, it was agreed to postpone approval of the minutes until later in the meeting.

Item 3 – Review and Approval of Minutes of the Previous Meeting #7, January 26<sup>th</sup>, 1998

There was insufficient time during the meeting to review the minutes. Al Morton submitted corrections and clarifications, which will be included in a revised version to be posted on the G-2.1.6 web site.

Item 4 – Matters Arising from the Minutes

As minutes were not reviewed, this item was not addressed.

Item 5 – Report of recent Meetings of ITU pertinent to video compression - Arthur Webster, David Fibush & other participants

Arthur Webster reported on the ITU Study Group 12 (SG12) meeting. He distributed a summary and meeting reports for Questions 10 and 11 from SG12. Refer to *Summary and meeting reports for Questions 10 and 11 from ITU-T-SG12: Objective and subjective methods for evaluating audiovisual quality in multimedia services*, Arthur Webster, NTIA/ITS, March 16, 1998. (IEEE Doc. G-2.1.6/76)

Arthur Webster said there was concern that CCETT felt it was taking on more of the work than it could handle free. Since people probably won't want to pay, we will either have to find subjective testing labs that will do it free or we will have to reduce the testing. Options suggested were to do the lower bit rate tests first or use a sparse matrix design with half as many scenes. The merits of paying for testing were discussed. It was suggested that "in-kind" contributions could be one way to cover costs.

Arthur Webster pointed out that there were some changes in the video quality table. Examples were added for cable, broadcast and video-conferencing.

The timetable for VQEG / ITU work was discussed. May 1, 1998 is the deadline to submit a letter of intent and a description of the algorithm for measurement of objective video quality. June 1 was the date for submitting code, but it was felt this date would probably shift to July 1. The subjective testing would take place during the summer, with data analysis finished in late October. There was concern that resources could be a problem, but David Fibush noted that the plan passed out at the January meeting hasn't really changed, except for the video table and audio table.

#### *5.1 Further Discussion and Recommendations from the Subcommittee.*

In response to a question about the alignment striping at the beginning of the scenes, a discussion ensued concerning the need for alignment. The need for composite input and outputs on the measurement equipment was explained. The argument for including composite video was that analog is widely used today and at present, no set-top boxes have a digital output. It was noted that the test is designed to test video quality measurement algorithms. Alignment is not tested. The tests do not have to be in real time.

It was questioned whether people would be able to see degradation at the higher quality levels at 18 Mbps and 36 Mbps at a viewing distance of 5H. If the machine could detect the degradation, but humans couldn't it will distort the subjective data. There was also concern that if we don't measure below threshold, we won't know how well the measurements will work when we concatenate systems. It was suggested that the viewing distance be reduced or the bit rates set lower. There was some support for reducing the viewing distance to 4H or 3H for the higher bit rates. Although there was concern this may require two different objective models, it was noted this could be handled through different settings, not different measurement methods. As this topic had been discussed before, the decision was made to evaluate it in the future.

#### Item 6 - Report of Task Force on Timetable, Chair, Leon Stanger

Leon Stanger reported there had been no activity since the last meeting. He felt there was nothing meaningful to do and offered to give the chair to someone else.

#### *6.1 Further Discussion and Action.*

Alan Godber indicated he wanted a time scale to see how we were progressing and to set goals for different points on the time scale. At present the only time scale is the work for the ITU group.

It was noted that a timetable might not be appropriate, since we can't put a schedule on invention of technology by a committee.

The issue was put in abeyance until the next meeting, when it will be reviewed to see if we need to take further action.

#### Item 7 - Report of Task Force on Compression Measurements Information Gathering - Chair, Bill Zou.

Alan Godber reported for Bill Zou. He noted that Dr. Andrew Watson and Dr. Christian J. van den Branden Lambrecht were here because of Zou's work. John Beerends has been contacted but no reply had been received. KDD sent a document, which was distributed at the last meeting. NHK was contacted. No reply has been received.

#### Item 8 - Presentation by Dr. Christian J. van den Branden Lambrecht, of the Imaging Technology Department, Hewlett Packard, on his proposals for Video Compression Measurement Techniques.

Dr. Christian J. van den Branden Lambrecht from HP laboratories delivered his presentation. Refer to *Presentation to IEEE G-2.1.6, Vision Models and Metrics for Video*, Christian J. van den Branden Lambrecht,

Hewlett-Packard Laboratories, March 16, 1998. (IEEE Doc. G-2.1.6/77) It is available in Adobe Acrobat format from the G-2.1.6 web site at <http://grouper.ieee.org/groups/videocomp/vdb9803.pdf> (size 1.22M).

### 8.1 Question and Answer Period and Discussion

The subjective tests described in the presentation used only a small patch of the total picture area. How do you extrapolate from the patch to the full screen of video?

Are artifacts introduced one at a time? The scenes were run through an encoder but were not processed one artifact at a time. It would be possible to do this, however. The test gear was built so that it can stress one aspect more than another, so, in effect it tested one artifact more.

Are the scenes artificial? Yes, for now. Eventually more than half will be natural scenes. Although the scenes are only 1.2 seconds long, they could include a transition if it was arranged for cut to occur at the right time in the scene.

Christian van den Branden Lambrecht was asked whether his studies looked at errors in motion estimation. He answered that he looked at the accuracy of the motion vector and had designed a motion estimation algorithm for the BBC encoder.

### Item 9 - Presentation by Dr. Andrew B. Watson, Senior Scientist for Vision Research, Flight Management and Human Factors Division, NASA Ames Research Center, on his Experience with Vision Research and Video Compression.

Dr. Watson described his work on NASA's Digital Video Quality metric. He noted that all current visual quality models are actually fidelity models. Fidelity models measure visible differences between a reference image and a test image. Thus, all video quality metrics are based on models of human vision. The metrics vary only in the accuracy of the model. The DCTune metric creates an error map based on the differences from a reference image after the image is passed through a visual transform. The accuracy of the model was tested by comparing subjective measurement results with the DCTune metric results. Much of the information presented in this paper is available on the NASA Ames Research Center Vision Science web page, <http://www.visionscience.com/VisionScience.html>.

The NASA Digital Video Quality Metric incorporates a spatial temporal chromatic contrast sensitivity function, is computationally very efficient, has a nested structure that allows tradeoffs between speed and accuracy and exploits existing hardware and software for video process such as DCT.

Dr. Watson recommended the VQEG tests vary the viewing distance. It is an important determinant of visual quality and it varies widely in the real world. In response to a question asking why one viewing distance wouldn't work, since at one viewing distance a device will measure one JND, Dr. Watson replied that JNDs would not be proportional to viewing distance. We need to measure them and will be interested in different methods. Metrics should work over the full range and we should be able to validate them over the full range.

He also recommended omitting the analog HRCs from the VQEG tests, noting that analog artifacts raise issues not relevant to digital video. They require markers, exclude in-service testing and their application is unclear. He acknowledged that there was no digital output on set-top boxes, but asked why we wanted to measure A/D converters. Comments from the subcommittee included the observation that we need to provide a measurement system that includes the analog portions of the system because they will be used. They exist in current broadcast plants and will continue to exist for one to ten years or more. Another comment was that the system needed to remove the analog artifacts, identify what it did and report on it. It was also suggested that there is a research reason to perturb space here and there and see what happens. Dr. Watson answered that if the number of HRCs was reduced, the number of viewing distances could be increased.

Dr. Watson also recommended the measurement of thresholds, which are an absolute measure, rather than some relative measure. He noted that observers have multi-dimensional impairment scales and will rate them differently. We can never hope to have human observers produce the same result. Although the results can be averaged to produce a mean, they remain limited by their multi-dimensional nature. He envisioned a real application that continuously tells that the quality degradation is just perceptible. For this, the threshold is the

most important data point. The DSCQS test method planned for the VQEG tests doesn't produce that data point.

### *9.1 Question and Answer Period and Discussion*

This comment led to a discussion of multiple JNDs. Several members had a problem with the concept of multiple JNDs in general and the additivity of JNDs in particular. The main question was how do you subjectively measure multiple JNDs. It was suggested you could measure one JND and use that as a reference for a second study to determine two JNDs. It was also agreed that human measurement data probably wouldn't show this kind of structure. One example given was the perception of rain. If you feel one drop of rain its raining. What happens if you feel four drops or ten drops? Some member felt that even if the multiple JNDs couldn't be defined precisely, there was a still need to define a scale. David Fibush pointed out that the Tektronix box did not use JNDs as the unit of measure. There was another comment that multiple JNDs, whether used for color charts or picture quality, are a fiction. They were extrapolated from the measurement of a single JND. In some ways, they work and in other ways, they don't.

There was agreement that it would be a benefit if a unit could measure "multiple JNDs". The problem arises when comparing the linearity of measurement of multiple JNDs with subjective measurements. An objective measurement must take into account that the human subject will be more sensitive to some degradation and consider this when producing a rating. The temporal filter, the spatial contrast sensitivity function, dramatic contrast sensitivity and such are all ways of weighting errors differently when calculating one JND.

Dr. Watson was asked that although he recommended eliminating composite video from the testing, since the NASA metric is a visual metric, did he know that it wouldn't work with the artifacts from composite video. He answered that if the images are aligned and the gain differences removed, the metric should work with composite artifacts.

Dr. Watson claimed his metric was fast. He was asked to clarify this. One example was given where fast was defined as being able to process a few seconds of video in a few hours. Dr. Watson said his metric should be able to run in real-time in a hardware implementation. David Fibush commented that while some models might run slow in software, they could be real time in hardware. Therefore, VQEG was looking for the strongest base algorithms first.

Dr. Watson he felt it was unlikely VQEG would come up with an algorithm that was a clear winner in the subjective testing. The discussion returned to JNDs, with the comment that while we could determine what one JND was, the problem was defining a scale beyond one JND. David Fibush summarized the problem by giving an example: What if one volt only meant this had a higher potential than that - nothing else. How would we define two volts?

The final comment was related to the recommendation that the subjective testing be based on threshold rather than a quality scale. While agreeing with need to do some tests using this method, the comment was that this was not what we were trying to do now. This caused some confusion, since both methods measured video quality (actually, fidelity). There was concern the threshold would not cover the broad range of applications being considered. One method for reducing the "coarseness" of the measurement is the use of a variety of sequences covering different quality requirements. It was noted that there is a need for a scale that extends beyond one JND. NTSC is more than one JND!

Discussion continued. The DSCQS measurements selected by VQEG for subjective testing were compared with the threshold measurements recommended by Dr. Watson. There was general agreement the threshold methods had merit and there was also some concern expressed that VQEG may be on too fast a schedule, given the questions raised during the two presentations today.

### Item10 - Report of Task Force on "Defining A Unit of Measure & a Means of Calibration for Video Impairment", Leon Stanger.

Document distributed. Leon Stanger presented his report

### *10.1 Further discussion and Action*

Leon Stanger said the task force would go back into committee and see how much further they can pursue the methods outlined in the report. He wants to define a scale before defining an algorithm.

David Fibush reported that SMPTE was working on distribution of test scenes for compression testing. However, error concealment is a problem for D-1 and D-5 formats, particularly D-1. Data tapes will also be available, probably in different formats. Stanger agreed it would be nice to have pristine material for testing, but also wanted material that is degraded in different degrees. The VQEG database generated as result of its testing was suggested as being useful for this. The P.930, VIRIS work was also mentioned as potentially useful.

One concern was that if we build a series of scales using tapes with different impairments and different degrees of impairments, how do we mix the results. For example, if defect A and defect B are both considered a "1", how do you add them? Stanger replied that this is the reason his task force limited it to singular defects.

Another concern was that a subjective component was involved in making the comparisons needed to use the reference tapes. That is a difficult step in the process. Experts would be needed for the comparison. Stanger, disagreed, describing the process as comparing the reference tapes with the processed material and estimating how it looks. There was another comment that the comparison would be similar to measuring signal to noise on a scope by adding noise to match the existing noise. That method has been shown to work.

Discussion continued on the method for measuring the amount of degradation for creating and calibrating the reference tapes. Given moving images, people will look at different parts of the stream. Do we have enough viewers? It was noted that an objective metric has an inherent advantage here. With 1000 measurements, a reasonably good machine should give similar numbers. For subjective measurements, it was suggested we choose the viewer who gives the image the worst rating.

Mention of the problem of variability among viewers making subjective measurements led to a discussion on test techniques. The length of the scenes was discussed. Should it be ten seconds? People tend to remember what they saw last. Recency is an issue. If it is built into the metric it might help it win the contest, but will it work in the real world? It was suggested the scenes be five or ten seconds in length, but that they repeat. In response, it was noted that there are learning effects -- you detect more if you look a scene a hundred times. Is this desired? People who make business decisions will look at things with an expert viewers eye. Alan Godber reported that at ATTC, using ten second images, and at CRC, expert and non-expert viewers saw the same things and described the same artifacts. Some members commented that expert viewers were needed. They noted that with DVD, people are able to see the same scene over and over. Another comment was that non-expert viewers become expert over time. If we use expert viewers now, we narrow the uncertainty. However, others felt that tests were still needed with non-experts.

Members were especially interested in finding test material that was subjectively rated. It was noted that it is difficult to obtain rights to use the material. Much of it is in composite video formats. David Fibush reported that SMPTE was working on distributing the raw test materials and that the VQEG material may be able to be put into the public domain. The 4:2:2 database is available but it is not to be made public. In response to a question as to whether VQEG could make a lack of copyright or other constraints on test material a requirement for submission, Arthur Webster said he will look into it. He remarked that while material is available for research, copyright holders don't want a piece of their movies degraded and shown on a web site. A suggestion was made that we ask ATTC to make material available at a nominal cost. Alan Godber requested a copy of any existing subjectively-rated test material for use by the committee.

Chairman Godber asked if this task force should continue. There was a consensus that it should. It was noted that defining a single unit for a quality scale would be a big achievement. Leon Stanger said the task force would try to pin down a few things.

#### Item 11 - Where Measurements Take Place in a Broadcast Chain - Leon Stanger.

This item was deferred.

#### Item 12 - Further Discussion of Compression Measurement Methodologies.

This item was deferred

Item 13 - Any Other Business.

Alan Godber reported that SMPTE had no problem with us posting papers from the SMPTE meeting for use by the committee provided we obtain permission from the authors of the papers.

Item 14 - Date(s) of Future Meeting(s).

There was no desire to hold a meeting before the next T1A1 meeting in July. It was decided that the next meeting would be one day before the T1A1 meeting at the same location, which was undetermined.

The committee offered thanks to T1A1 and NTIA / ITS for use of the facilities.

The meeting was adjourned at 6:55 PM MST.

Submitted by:

H. Douglas Lung

Secretary

APPENDIX "A"

List of Documents Distributed

16 March 1998

*Revised Draft Agenda - IEEE Compression and Processing Subcommittee G-2.1.6, Eighth Meeting, Monday, March 16, 1998, Alan Godber, Chairman, 13 March 1998.*

*Draft Meeting Record, G-2.1.6, Compression and Processing Subcommittee, Meeting #7, January 26, 1998, Advanced Television Technology Center, 1330 Braddock Place, Suite 200, Alexandria, Virginia, Doug Lung, Secretary, Doc. G-2.1.6/75, 8 March, 1998, Revised 16 March 1998.*

*Summary and meeting reports for Questions 10 and 11 from ITU-T-SG12: Objective and subjective methods for evaluating audiovisual quality in multimedia services, Arthur Webster, NTIA/ITS, March 16, 1998, Doc. G-2.1.6/76, 16 March 1998.*

*Presentation to IEEE G-2.1.6, Vision Models and Metrics for Video, Christian J. van den Branden Lambrecht, Hewlett-Packard Laboratories, March 16, 1998, Doc. G-2.1.6/77, 16 March 1998.*

*Progress Report of Task Force to Define a Unit of Measure and Means of Calibration for Video Quality Analysis, Leon Stanger, 11 March 1998, Doc. G-2.1.6/78, 16 March 1998.*

APPENDIX "B"  
ATTENDANCE RECORD  
16 March 1998

<b>Name</b>	<b>Affiliation</b>	<b>Telephone</b>	<b>Fax</b>	<b>E-mail</b>
Chairman: Alan Godber	Consultant	(908) 846-4476	(908) 846-4476	<a href="mailto:agodber@mail.idt.net">agodber@mail.idt.net</a>
Secretary: Doug Lung	Telemundo	(305) 884-9664	(305) 884-9661	<a href="mailto:dlung@transmitter.com">dlung@transmitter.com</a>
Lorence Brown	Ameritech	(847) 248-4379	(847) 248-6746	<a href="mailto:lorence.brown@ameritech.com">lorence.brown@ameritech.com</a>
Emel C. Celi	Sprint / T1A1	(650) 375-4506	(650) 375-4599	<a href="mailto:mceli@sprintlabs.com">mceli@sprintlabs.com</a>
David Fibush	Tektronix	(503) 627-6289	(503) 627-1707	<a href="mailto:davef@tv.tv.tek.com">davef@tv.tv.tek.com</a>
Paul Jones	Kodak	(716) 477-8048		<a href="mailto:pjones@kodak.com">pjones@kodak.com</a>
John Libert	NIST	(301) 975-3828	(301) 926-3534	<a href="mailto:libert@eeel.nist.gov">libert@eeel.nist.gov</a>
Al Morton	AT&T	(908) 949-2499	(908) 949-1652	<a href="mailto:acmorton@att.com">acmorton@att.com</a>
Margaret H. Pinson	NTIA	(303) 497-3579	(303) 497-5323	<a href="mailto:margaret@its.bldrdoc.gov">margaret@its.bldrdoc.gov</a>
James R. Redford	NBC	(212) 664-5222	(212) 246-3650	<a href="mailto:rick.redford@nbc.com">rick.redford@nbc.com</a>
Leon Stanger	DirectTV	(310) 726-4676	(310) 726-4535	<a href="mailto:LStanger@compuserve.com">LStanger@compuserve.com</a>
Dick Streeter	Representing CBS	(908) 791-9876	(908) 791-9878	<a href="mailto:rstreeter@msn.com">rstreeter@msn.com</a>
Pat Tweedy	GTE Labs	(617) 466-2661	(617) 466-2650	<a href="mailto:ept0@gte.com">ept0@gte.com</a>
C. van den Branden	HP	(650) 857-7658	(650) 857-4691	<a href="mailto:volb@hpl.hp.com">volb@hpl.hp.com</a>
Andrew Watson	NASA	(650) 604 5419	(650) 604-0255	<a href="mailto:abwatson@mail.arc.nasa.gov">abwatson@mail.arc.nasa.gov</a>
Arthur Webster	NTIA/ITS	(303) 497-3567	(303) 497-5323	<a href="mailto:webster@its.bldrdoc.gov">webster@its.bldrdoc.gov</a>
Stephen Wolf	NTIA	(303) 497-3771	(303) 497-5323	<a href="mailto:steve@its.bldrdoc.gov">steve@its.bldrdoc.gov</a>

